

# Inconsistency of Measure-Theoretic Probability

Guang-Liang Li \*

Victor O.K. Li

Department of Electrical and Electronic Engineering

The University of Hong Kong

{glli,vli}@eee.hku.hk

December 18, 2014

## Abstract

We reveal a contradiction in measure-theoretic probability. The contradiction is an “equation”  $1/2 = 0$  with its two sides representing probabilities. Unlike known paradoxes in mathematics, the revealed contradiction cannot be explained away and actually indicates that measure-theoretic probability is inconsistent. Appearing only in the theory, the contradiction does not exist in the physical world. So practical applications of measure-theoretic probability will not be affected by the inconsistency as long as “ideal events” in the theory (which will never occur physically) are not mistaken for real events in the physical world. Nevertheless, the inconsistency must be resolved. Constructive mathematics can avoid such inconsistency. There is no contradiction reported in constructive mathematics.

## 1 Introduction

Known paradoxes in mathematics are either considered issues of mathematical logic [1], or treated as deducible statements which are counterintuitive but true [6, 10], where “true” implies “without contradiction”. Treating deducible statements as results without contradiction is not grounded on mathematical justification. Such treatment is usually based on a popular belief, i.e., mathematics is free from contradiction as it is. The belief is an issue of mathematical philosophy, which is not our concern. In this paper, we show that a contradictory “equation”  $1/2 = 0$  is deducible in measure-theoretic probability, where the two sides of the “equation” represent probabilities. Unlike the paradoxes discussed in [6, 10], the contradictory “equation” cannot be explained away and actually indicates that measure-theoretic probability is inconsistent.

After a brief discussion about why measure-theoretic probability is very successful and yet inconsistent (Section 2), we review some key notions relevant to revealing the inconsistency, such as inadequacy of the  $\sigma$ -algebra generated by coordinate variables of stochastic processes (Section 3), weak convergence and tightness of probability measures on the  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R})$  of subsets of the real line  $\mathbb{R}$  (Section 4), weak convergence of probability measures on the  $\sigma$ -algebra  $\mathcal{B}(\overline{\mathbb{R}})$  of subsets of the extended real line  $\overline{\mathbb{R}}$  (Section 5), and the difference between probabilities and numbers in the closed unit interval  $[0, 1]$  (Section 6), in particular, the difference between limiting probabilities and limits of sequences in  $[0, 1]$  (Section 7). Based on the above preparation, we demonstrate the deducibility of the contradictory “equation”  $1/2 = 0$  (Section 8). We conclude in Section 9.

---

\*Corresponding author (alternative e-mail address: guanglli@hotmail.com). 2010 *Mathematics Subject Classification*: Primary 60A05.

## 2 Measure-Theoretic Probability: Success and Inconsistency

For measure-theoretic probability, the ultimate criterion of success is whether the probability of a real event (i.e., an event in the physical world) predicted by the theory can be verified empirically. By this criterion, measure-theoretic probability is very successful. And yet, measure-theoretic probability is inconsistent, as we shall demonstrate. How can measure-theoretic probability be both successful and inconsistent? A brief answer is as follows. As a consequence of “ideal events”, the inconsistency of measure-theoretic probability appears only in the theory, but not in the physical world. Expressed as an infinite union or intersection of different events of positive probability, an “ideal event” is a result of infinitely repeated or infinitely fine operations.

**Example 1** An absolutely accurate measurement of a physical quantity with its relative magnitude represented by an irrational number is an ideal event. Corresponding to picking up a point from a set with the cardinality of the continuum, such measurement requires infinitely fine operations, and hence is not physically realizable.  $\square$

The criterion of success of measure-theoretic probability implies a constraint: Since “ideal events” do not exist physically, or since it is impossible to verify the existence of “ideal events” in the physical world, physical existence of theoretical properties derived by manipulating probabilities of “ideal events” should not be inferred from experiments. This constraint serves as a “firewall” of measure-theoretic probability. When the constraint is satisfied, it is safe to apply measure-theoretic probability to solve problems in the real world. In practice, however, violating the constraint can cause confusion of real and ideal events. A typical violation is to use a finite set of experimental data to “prove” or “verify” physical existence of theoretical properties obtained by calculating probabilities of “ideal events”.

## 3 Inadequacy of $\sigma$ -Algebra Generated by Coordinate Variables

A real-valued random variable is usually described by the probability measure (i.e., the distribution) induced by the random variable on the real line  $\mathbb{R}$ . In general, a stochastic process is usually described by the probability measures (i.e., the finite-dimensional distributions) induced by the process in  $k$ -dimensional Euclidean spaces  $\mathbb{R}^k, k = 1, 2, \dots$ . The construction of a process in measure-theoretic probability begins with a given system of finite-dimensional distributions. However, the system of finite-dimensional distributions may not completely capture the properties of a process [3].

A stochastic process is a collection of real-valued random variables labeled by an index variable taking values from a set  $T$ . The points in the set  $T$  usually represent time. We consider only  $T = \{1, 2, \dots\}$ . Denote by  $\mathbb{R}^T$  the set of all real sequences. For each  $n \in T$ , define a mapping  $W_n : \mathbb{R}^T \rightarrow \mathbb{R}$  by

$$W_n(\omega) = \omega(n) = \omega_n.$$

For a given  $n \in T$ , the mapping  $W_n$  (also referred to as a coordinate function) maps a real sequence  $\omega \in \mathbb{R}^T$  to the  $n$ th coordinate  $\omega(n) = \omega_n$  of  $\omega$ . We may also write  $W_n(\omega)$  as  $W(n, \omega)$ . For a fixed  $\omega$ ,  $W(\cdot, \omega) = \omega(\cdot)$  is just  $\omega$  itself. Denote by  $\mathcal{B}(\mathbb{R}^T)$  the  $\sigma$ -algebra (also referred to as  $\sigma$ -field) generated by all the coordinate functions  $W_n, n \in T$ :  $\mathcal{B}(\mathbb{R}^T) = \sigma\{W_n : n \in T\}$ , which is the intersection of all the  $\sigma$ -algebras containing the sets of the form

$$\{\omega \in \mathbb{R}^T : W_n(\omega) \in H\} = \{\omega \in \mathbb{R}^T : \omega_n \in H\}$$

for  $n \in T$  and  $H \in \mathcal{B}(\mathbb{R})$ . In other words,  $\mathcal{B}(\mathbb{R}^T)$  is the “smallest” class of the sets which is closed under the finite and countable set-theoretic operations.

Let  $\mathbb{R}^T$  be a sample space represented by  $\Omega$ , and let a probability measure be defined on  $\mathcal{B}(\mathbb{R}^T)$ . Now  $\omega \in \Omega = \mathbb{R}^T$  represents a sample point, and  $W_n$  can be interpreted as the  $n$ th coordinate variable  $U_n$  of a random sequence  $(U_n)_{n \geq 1}$ . A realization of  $(U_n)_{n \geq 1}$  at a given sample point  $\omega$  is a real sequence  $(U_n(\omega))_{n \geq 1}$ , which can also be expressed as  $U(\cdot, \omega)$  and is just  $\omega$  itself.

If  $(U_n)_{n \geq 1}$  is a non-decreasing random sequence,  $U_n(\omega)$  may tend to infinity as  $n \rightarrow \infty$  at some  $\omega$ . Since  $\{\infty\} \notin \mathcal{B}(\mathbb{R}^T)$ , a probability measure on the measurable space  $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$  is not useful to describe the limiting behavior of  $(U_n)_{n \geq 1}$ . Clearly, to define the limiting random variable of  $(U_n)_{n \geq 1}$  on a probability space, the limiting random variable should be measurable with respect to the corresponding probability measure, which requires the values of the limiting random variable to form a measurable set on the probability space. Fortunately, the probability measures  $\rho_n$  induced by  $U_n$  constitute a sequence  $(\rho_n)_{n \geq 1}$  such that, corresponding to  $(U_n)_{n \geq 1}$  and  $(\rho_n)_{n \geq 1}$ , a limiting random variable  $U$  and a limiting probability measure  $\rho$  induced by  $U$  exist on the extended real line  $\overline{\mathbb{R}}$  and the corresponding  $\sigma$ -algebra  $\mathcal{B}(\overline{\mathbb{R}})$ , respectively. For a convergent sequence of probability measures, the corresponding random variables can be on different probability spaces [4]. Consequently, convergence of probability measures does not require the involved random variables to form a convergent sequence on the same probability space. Convergence of probability measures on  $\mathcal{B}(\overline{\mathbb{R}})$  is a generalization of convergence of probability measures on  $\mathcal{B}(\mathbb{R})$ . If a sequence of probability measures converges on  $\mathcal{B}(\mathbb{R})$ , the sequence is necessarily tight on  $\mathcal{B}(\mathbb{R})$ . We shall review these notions in the next two sections.

## 4 Weak Convergence and Tightness of Probability Measures on $\mathcal{B}(\mathbb{R})$

Let  $(\rho_n)_{n \geq 1}$  be a sequence of probability measures on  $\mathcal{B}(\mathbb{R})$ . The probability measures  $\rho_n$  are distributions induced by real-valued random variables  $U_n$ , such that  $U_n$  converge in distribution to a limiting random variable  $U$  as  $n \rightarrow \infty$  (notation  $U_n \Rightarrow U$ ). Equivalently, the probability measures  $\rho_n$  converge weakly to a limiting probability measure  $\rho$  induced by  $U$  (notation  $\rho_n \Rightarrow \rho$ ). If  $(U_n)_{n \geq 1}$  is non-decreasing,  $U$  may take  $\infty$  as its value with positive probability, i.e.,  $\rho$  may assign positive probability to  $\{\infty\}$ . One way to define convergence in distribution of a sequence of real random variables is by convergence of the corresponding distribution functions at each continuous point of the limiting distribution function, which may not necessarily be a distribution function of a real random variable [7]. This definition has some disadvantages, for example, see [8].

An equivalent definition of weak convergence of probability measures (i.e., convergence in distribution of the corresponding random variables) is as follows [4, 8]. Suppose  $\mathbb{R}$  is the common range for all the random variables involved. The sequence  $(\rho_n)_{n \geq 1}$  of probability measures induced by the random variables converges weakly to a probability measure  $\rho$ , if and only if for each  $f \in \mathbb{C}_b(\mathbb{R})$ ,

$$\int_{\mathbb{R}} f d\rho_n \rightarrow \int_{\mathbb{R}} f d\rho, \quad n \rightarrow \infty$$

where  $\mathbb{C}_b(\mathbb{R})$  is the class of bounded, continuous functions on  $\mathbb{R}$ . Clearly, such convergence is not only a property of the probability measures  $\rho_n$  or corresponding random variables  $U_n$ , but also relevant to the common range space of the involved random variables on which weak convergence is defined. If the common range space is defined to be  $\mathbb{R}$ , and if the limiting random variable is not real-valued, then weak convergence or convergence in distribution as defined above does not hold, even if a limiting random variable and the corresponding probability measure exist on  $\overline{\mathbb{R}}$  and  $\mathcal{B}(\overline{\mathbb{R}})$ , respectively.

**Example 2** Let  $(U_n)_{n \geq 1}$  be a sequence of degenerate random variables such that  $U_n = n$ . The distribution, i.e., the probability measure  $\rho_n$  induced by  $U_n$  is the Dirac measure concentrated at  $\{n\}$ . As

$n \rightarrow \infty$ ,  $U_n \Rightarrow U = \infty \in \overline{\mathbb{R}}$  and  $\rho_n \Rightarrow \rho$  where  $\rho$  is the Dirac measure concentrated on  $\{\infty\} \in \mathcal{B}(\overline{\mathbb{R}})$ . But  $(U_n)_{n \geq 1}$  does not converge on  $\mathbb{R}$ , and  $(\rho_n)_{n \geq 1}$  does not converge weakly on  $\mathcal{B}(\mathbb{R})$ .  $\square$

On the one hand, as illustrated by the above example, for a sequence of probability measures, mass may “escape to infinity” as  $n \rightarrow \infty$ , and meanwhile, in an extended sense, the notion of limiting probability measure remains meaningful. In particular, for a non-decreasing random sequence such as  $(U_n)_{n \geq 1}$  in Example 2, a unique limiting random variable exists, which may not necessarily be real-valued. By Helly’s selection theorem (e.g., see [3]), the limiting random variable has unique distribution function and probability measure, such that the limiting distribution function may have a lower bound larger than 0 or an upper bound less than 1, and the limiting probability measure  $\rho$  is on  $\mathcal{B}(\overline{\mathbb{R}})$  and may be a subprobability measure when restricted to  $\mathcal{B}(\mathbb{R})$ , i.e.,  $\rho(\mathbb{R}) < 1$ . The cause of  $\rho(\mathbb{R}) < 1$  is mass “escaping to infinity”.

On the other hand, probability measures defined on  $\mathcal{B}(\mathbb{R})$  are particularly important for measure-theoretic probability and its applications. To prevent mass from “escaping to infinity”, a condition must be imposed on a sequence of probability measures before we consider whether the sequence converges to a limiting probability measure on  $\mathcal{B}(\mathbb{R})$ .

Tightness of probability measures on  $\mathcal{B}(\mathbb{R})$  is the condition to prevent mass from “escaping to infinity”. If a sequence of probability measures  $(\rho_n)_{n \geq 1}$  is tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , then for each  $\epsilon > 0$ , there is an interval  $I_\epsilon \subset \mathbb{R}$ , such that  $\sup_n \rho_n(I_\epsilon^c) < \epsilon$  where  $I_\epsilon^c = \mathbb{R} \setminus I_\epsilon$ . If such a tight sequence of probability measures  $(\rho_n)_{n \geq 1}$  converges weakly to  $\rho$ , then the weak limit  $\rho$  is also a probability measure on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Given  $f \in \mathbb{C}_b(\mathbb{R})$ , even if  $f(x)$  does not converge as  $x \rightarrow \pm\infty$ , so long as  $(\rho_n)_{n \geq 1}$  is tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and  $\rho_n \Rightarrow \rho$ , we have  $\rho(\{\pm\infty\}) = 0$  and  $\int_{\mathbb{R}} f d\rho$  exists. Suppose  $\rho_n \Rightarrow \rho$ . If not only all  $\rho_n$  but also  $\rho$  are required to be probability measures on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , then  $(\rho_n)_{n \geq 1}$  is necessarily tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

**Example 3** Let  $\lambda, \lambda_n, n = 1, 2, \dots$  be probability measures on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , such that  $\lambda, \lambda_1, \lambda_2, \dots$  are distributions of i.i.d. random variables  $X, X_1, X_2, \dots$ , respectively. The probability mass functions of  $X$  and  $X_n$  are the same:

$$\lambda(\{j\}) = \lambda_n(\{j\}) = \begin{cases} 1/2, & j = 0 \\ 1/2, & j = 1 \\ 0, & \text{otherwise.} \end{cases}$$

If we define  $\lambda(\{\pm\infty\}) = \lambda_n(\{\pm\infty\}) = 0$ , then  $\lambda, \lambda_n, n = 1, 2, \dots$  are also probability measures on  $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$  trivially. The sequence  $(\lambda_n)_{n \geq 1}$  is tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , and converges weakly to  $\lambda$ . Equivalently,  $X_n$  converge in distribution to  $X$ .  $\square$

**Example 4** Consider a random sequence  $(Z_n)_{n \geq 1}$ . Denote by  $\mu_n$  the probability measure (i.e., the distribution) induced by  $Z_n$ , and let the probability mass function of  $Z_n$  be

$$\mu_n(\{j\}) = \begin{cases} 2^{-(n-j+1)}, & 1 \leq j < n \\ 2^{-1} + 2^{-n}, & j = n \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

For each  $n$ ,  $Z_{n+1}$  takes larger values with larger probabilities. The values of  $Z_{n+1}$  include  $n+1$  and all the values of  $Z_n$ .

The sequence  $(\mu_n)_{n \geq 1}$  is not tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . To see this, consider  $\epsilon = 2^{-1}$ . By (1), for all  $n$ ,  $\mu_n(\{n\}) = 2^{-1} + 2^{-n} > 2^{-1}$ . In particular, for all real number  $b > 0$  and all  $n > N_b$ ,  $\mu_n(\{n\}) > 2^{-1}$ , where  $N_b = \lceil b \rceil$ , and  $\lceil b \rceil$  is the smallest integer larger than or equal to  $b$ . Consequently, for each interval

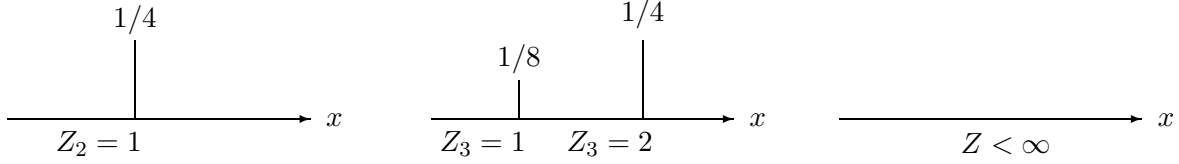


Figure 1. For any fixed  $k \geq 0$ , mass at  $\{n - k\}$  “escapes to infinity” as  $n \rightarrow \infty$ , so  $\mu(\mathbb{R}) = 0$ .

$I_b = (-\infty, b)$ ,  $\sup_n \mu_n(I_b^c) > 2^{-1}$ . The probability mass function (1) explains not only why  $(\mu_n)_{n \geq 1}$  fails to be tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  but also where all the mass goes as  $n \rightarrow \infty$ . For a fixed  $j$ ,  $\lim_n \mu_n(\{j\}) = 0$ . To see what happens when  $j$  increases with  $n$ , write  $j = n - k$ , where  $0 \leq k < n$  is fixed. We then have  $\mu_n(\{n - k\}) = 1/2^{k+1}$  for  $k > 0$  and  $\mu_n(\{n\}) = 2^{-1} + 2^{-n}$ . Since  $(n - k) \rightarrow \infty$  as  $n \rightarrow \infty$ ,  $1/2^{k+1}, k = 1, 2, \dots$  constitute one half of the mass that “escapes to infinity”. The other half of the mass at  $\{\infty\}$  in the limit as  $n \rightarrow \infty$  comes from  $\mu_n(\{n\})$ . Let a vertical line segment of length  $\mu_n(\{j\})$  at  $j$  represent the mass assigned to  $\{j\}$  by  $\mu_n$ . As  $n \rightarrow \infty$ , the length of the segment at a fixed  $j$  diminishes towards 0, the segment at  $j = n - k$  with a fixed  $0 < k < n$  maintains an unchanged length  $2^{-(k+1)}$ , and the length of the segment at  $n$  is  $2^{-1} + 2^{-n}$ , which decreases gradually to  $1/2$ . Figure 1 illustrates the case for  $k > 0$  with  $n = 2$  and  $3$ . The total “escaped” mass then is  $\sum_{k \geq 0} 1/2^{k+1} = 1$ . Actually, by Helly’s theorem, each subsequence of  $(\mu_n)_{n \geq 1}$  converges to a subprobability measure  $\mu$  on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  such that  $\mu(\mathbb{R}) = 0$ , for  $\mu(\mathbb{R}) > 0$  contradicts (1). So  $\mu_n \Rightarrow \mu$  where  $\mu$  is the Dirac measure on  $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$  with  $\mu(\{\infty\}) = 1$ . Equivalently,  $Z_n \Rightarrow Z$ , where  $Z$  is the extended random variable corresponding to  $\mu$ . When restricted to  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ,  $\mu$  is a zero measure.  $\square$

As we shall see later,  $(\lambda_n)_{n \geq 1}$  in Example 3 has an intrinsic connection to  $(\mu_n)_{n \geq 1}$  in Example 4, though the former is tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and the latter is not.

## 5 Weak Convergence of Probability Measures on $\mathcal{B}(\overline{\mathbb{R}})$

The definition of  $\rho_n \Rightarrow \rho$  can be generalized such that, for each  $n$ ,  $\rho_n$  is still a probability measure on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , but  $(\rho_n)_{n \geq 1}$  may not be tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , and  $\rho$  is a probability measure on  $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$  with  $\rho(\{\pm\infty\}) > 0$  (see Example 2). In the above generalized definition of  $\rho_n \Rightarrow \rho$ ,  $\mathbb{C}_b(\mathbb{R})$  in the original definition is replaced by  $\mathbb{C}(\overline{\mathbb{R}})$ , the class of continuous functions on  $\overline{\mathbb{R}}$ . The reason for such replacement is shown by the example below. Let  $\mathbf{1}_A$  represent the indicator function, i.e.,

$$\mathbf{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & \text{otherwise.} \end{cases}$$

**Example 5** Consider  $(\rho_n)_{n \geq 1}$  such that  $\rho_n$  is the Dirac measure concentrated at  $\{n\}$  as in Example 2. The distribution function corresponding to  $\rho_n$  is  $\mathbf{1}_{[n, \infty)}(\cdot)$  with  $\mathbf{1}_{[n, \infty)}(x) \rightarrow 0$  as  $n \rightarrow \infty$  for  $\lim_n [n, \infty) = \cap_n [n, \infty) = \emptyset$ . By Helly’s selection theorem and the generalized definition of  $\rho_n \Rightarrow \rho$ ,  $\rho$  is the Dirac measure concentrated at  $\{\infty\}$ . However, for some  $f \in \mathbb{C}_b(\mathbb{R})$  such as  $f = \sin$ ,  $\int_{\mathbb{R}} f d\rho_n = f(n) = \sin(n)$ , which does not converge as  $n \rightarrow \infty$ .  $\square$

Let  $\overline{\mathbb{R}}$  be equipped with a topology induced by a fixed metric. A function from one metric space to another is called a homeomorphism if the function is continuous, one-to-one, onto, and its inverse

function is also continuous. If such a function exists, then the two metric spaces are said to be homeomorphic. With a homeomorphism from  $\overline{\mathbb{R}}$  onto  $[0, 1]$ ,  $\mathbb{C}(\overline{\mathbb{R}})$  can be identified with  $\mathbb{C}([0, 1])$ , the space of continuous functions on  $[0, 1]$ .

**Example 6** A homeomorphism  $h : \overline{\mathbb{R}} \rightarrow [0, 1]$  from  $\overline{\mathbb{R}}$  onto  $[0, 1]$  is

$$h(x) = \begin{cases} 0, & x = -\infty \\ 1 - \pi^{-1} \cot^{-1}(x), & x \in \mathbb{R} \\ 1, & x = \infty. \end{cases}$$

The inverse of  $h$  is  $h^{-1} : [0, 1] \rightarrow \overline{\mathbb{R}}$  given by

$$h^{-1}(y) = \begin{cases} -\infty, & y = 0 \\ \cot(\pi(1 - y)), & y \in (0, 1) \\ \infty, & y = 1. \end{cases}$$

A usual metric on  $\overline{\mathbb{R}}$  is  $d(x, y) = |\tan^{-1}(x) - \tan^{-1}(y)|$ .  $\square$

Each  $f \in \mathbb{C}([0, 1])$  has limits at the boundary points. Weak convergence of probability measures on  $\mathcal{B}(\overline{\mathbb{R}})$  can then be identified with weak convergence of probability measures on  $\mathcal{B}([0, 1])$  (the  $\sigma$ -algebra of subsets of  $[0, 1]$ ), and hence a limiting probability measure on  $\mathcal{B}([0, 1])$  which assigns mass to a boundary point of  $[0, 1]$  is identified with the corresponding limiting probability measure on  $\mathcal{B}(\overline{\mathbb{R}})$  which assigns the same amount of mass to  $\{\infty\}$ . Since the probability measures on  $\mathcal{B}([0, 1])$  may not necessarily be tight on  $\mathcal{B}((0, 1))$  (the  $\sigma$ -algebra of subsets of  $(0, 1)$ ), weak convergence of such probability measures on  $\mathcal{B}([0, 1])$  does not necessarily imply tightness of the corresponding probability measures on  $\mathcal{B}(\overline{\mathbb{R}})$ .

**Example 7** Consider  $(Z_n)_{n \geq 1}$  in Example 4 again. Write  $Z'_n = h(Z_n)$  where  $h$  is given in Example 6. Denote by  $\mu'_n$  the probability measure induced by  $Z'_n$ . As  $n \rightarrow \infty$ ,  $Z_n \Rightarrow Z$ . Accordingly  $Z'_n \Rightarrow Z'$ . Let  $\mu'$  be the probability measure induced by  $Z'$ . Now  $\mu'_n \Rightarrow \mu'$  on  $([0, 1], \mathcal{B}([0, 1]))$  corresponds to  $\mu_n \Rightarrow \mu$  on  $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ . Consequently,  $\mu'(\{1\}) = 1$  and  $\mu'((0, 1)) = 0$  correspond to  $\mu(\{\infty\}) = 1$  and  $\mu(\mathbb{R}) = 0$ , respectively. Although  $\overline{\mathbb{R}}$  and  $[0, 1]$  are homeomorphic,  $(\mu_n)_{n \geq 1}$  is still not tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Clearly, the correspondence between  $Z' = 1$  and  $Z = \infty$  does not imply that  $Z'$  and  $Z$  are equivalent in magnitude. The trivial tightness of  $(\mu'_n)_{n \geq 1}$  on  $([0, 1], \mathcal{B}([0, 1]))$  will not change the fact that  $(\mu_n)_{n \geq 1}$  is not tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .  $\square$

## 6 Probabilities and Real Numbers in $[0, 1]$

The notion of real numbers and the notion of probability can both be formulated axiomatically. However, the axioms for real numbers cannot formulate the notion of probability. As a set function in measure-theoretic probability formulated by the axioms of probability, a probability measure assigns a numerical probability (also referred to as probability mass) represented by a number in the closed unit interval  $[0, 1]$  to an event characterized by a measurable set on a probability space. The numerical probability quantifies the possibility of the event. In other words, associated with a measurable set on a specified probability space, a (numerical) probability describes quantitatively how likely an event will occur. Real numbers without the constraint imposed by the axioms of probability do not have such unique property of probabilities. Therefore, a number in  $[0, 1]$  is not necessarily a probability of an event on a given probability space. Probabilities must comply with the constraint imposed by the axioms of probability. Numbers in  $[0, 1]$  are not necessarily subject to such constraint. Consequently, although

the amount or relative magnitude of a probability is represented by a real number in  $[0, 1]$ , probabilities cannot be treated as such numbers.

Real numbers are an ordered field. In such a field, comparison of real numbers is defined by the axioms of order. “Possibility” is a quantifiable attribute of an event. Probabilities are compared as quantified possibilities. Since the amount of a probability is represented by a real number in  $[0, 1]$ , comparing probabilities is equivalent to comparing their amounts. In contrast, if a real number in  $[0, 1]$  fails to comply with the axioms of probability, then the number is not a quantified possibility of an event on the underlying probability space, and hence is not a probability. Comparing such a number with a probability is meaningless. Such comparison is not defined by the axioms for either probability or real numbers. Actually, relations on  $\mathbb{R}$  such as  $<$ ,  $>$ ,  $=$ ,  $\leq$  and  $\geq$  are all defined for and only for real numbers, where real numbers are merely “points” in the ordered field. No other interpretation is attached to such abstract points.

**Example 8** Unit probability is a unique quantity for indicating certainty or almost sure occurrence of an event. In contrast, the real number “1” is merely a point in an absolute space  $\mathbb{R}$  formulated as a set based on the axioms of real numbers without any other interpretation. “An event occurs with probability 1” is certainly a meaningful statement. However, if “with probability 1” is replaced by “with the real number 1”, the statement “an event occurs with the real number 1” is meaningless. Unit probability and the real number “1” are of the same amount, but not of the same kind, and hence not comparable.  $\square$

**Example 9** From information theory we can also see the difference between probabilities and numbers in  $[0, 1]$ . As shown in information theory, the notion of probability is closely related to the notion of “information”. Without the constraint imposed by the axioms of probability, numbers in  $[0, 1]$  have no such connection to “information”.  $\square$

Nevertheless, since the amount of a probability is represented by a real number in  $[0, 1]$ , according to the axioms of real numbers, we can still compare the amount of a probability with a real number. However, when making such comparison, we must keep in mind that a probability and its amount (i.e., its relative magnitude represented by a real number) are different.

An operation defined on a set is said to be “closed” with respect to a given property of some or all elements in the set, if every result of the operation is an element of the set and possesses the property. The values of probabilities are represented by numbers in  $[0, 1]$ . However, although operations defined for numbers in  $[0, 1]$  may be closed with respect to numbers, i.e., results of such operations are also numbers, the operations may not necessarily be closed with respect to probabilities. Any closed operation defined for probabilities must obey the axioms of probability. This requirement will not change the closeness of an operation defined for numbers. When a closed operation defined for numbers violates the axioms of probability, the result of the operation is not a probability, but is still a number.

**Example 10** A probability space has four elementary events represented by  $\omega_i$  with probabilities

$$P(\{\omega_i\}) = 1/8, i = 1, 2, 3, \text{ and } P(\{\omega_4\}) = 5/8.$$

For  $A = \{\omega_1, \omega_2\}$  and  $B = \{\omega_2, \omega_3\}$ , the condition for  $P(A)$  and  $P(B)$  to be additive is violated by  $A \cap B \neq \emptyset$ . Consequently  $P(A) + P(B) = 1/2$  is not a probability of any event on the probability space. Nevertheless, as the sum of two numbers,  $P(A) + P(B)$  is still a number.  $\square$

Let  $\rho$  and  $\rho_n, n = 1, 2, \dots$  be probability measures on a common measurable space  $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ , such that  $\rho_n \Rightarrow \rho$ , i.e.,  $\rho_n$  converges weakly to  $\rho$  as  $n \rightarrow \infty$ , where  $\rho(\{-\infty\}) = 0$  and  $\rho_n(\mathbb{R}) = 1$  for each  $n$ . Let

$\alpha$  and  $\alpha_n, n = 1, 2, \dots$  be measurable sets in  $\mathcal{B}(\overline{\mathbb{R}})$ , such that  $\alpha = \lim_n \alpha_n, n \rightarrow \infty, \alpha_n \in \mathcal{B}(\mathbb{R}) \subset \mathcal{B}(\overline{\mathbb{R}})$ , and  $\rho_n(\alpha_n) < 1$  for all  $n$ . Suppose  $\lim_n \rho_n(\alpha_n)$  exists. If  $\alpha = \mathbb{R}$ , and if  $\lim_n \rho_n(\alpha_n)$  is treated as the limit of a real sequence, i.e., if  $\lim_n \rho_n(\alpha_n)$  is interpreted as  $\lim_n [\rho_n(\alpha_n)]$ , then the result of such treatment may not necessarily be a probability.

**Example 11** Let  $\rho_n$  be  $\mu_n$  as given in Example 4. Suppose  $\alpha_n = I(x_n) = (-\infty, x_n]$ . Consider  $x_n = n - 1$ .

$$\rho_n(\alpha_n) = \mu_n(I(n - 1)) = \sum_{j=1}^{n-1} \mu_n(\{j\}) = \sum_{j=1}^{n-1} \left( \frac{1}{2^{n-j+1}} \right) = \frac{1}{2} - \frac{1}{2^n}.$$

Suppose  $\lim_n \mu_n(I(n - 1))$  is treated as

$$\lim_{n \rightarrow \infty} [\mu_n(I(n - 1))] = \lim_{n \rightarrow \infty} \left( \frac{1}{2} - \frac{1}{2^n} \right) = \frac{1}{2}.$$

However, no set in  $\mathcal{B}(\overline{\mathbb{R}})$  is assigned one half of unit mass by any of the involved probability measures  $\mu, \mu_1, \mu_2, \dots$ . So  $\lim_n [\mu_n(I(n - 1))] = 1/2$  is not a probability, and the operation of taking limit is not closed with respect to probabilities. Nevertheless,  $\lim_n [\mu_n(I(n - 1))] = 1/2$  is still the limit of a real sequence. The operation of taking limit remains to be closed with respect to numbers for any convergent sequence in  $[0, 1]$ .  $\square$

## 7 Limiting Probabilities on $\mathcal{B}(\mathbb{R})$ and Limits of Real Sequences in $[0, 1]$

Limiting probabilities are not necessarily limits of probabilities. The former must be probabilities. But the latter may or may not be probabilities. A limiting probability is determined by the corresponding limiting measure and event. The event may not necessarily be a continuity set of the limiting measure [4]. Suppose  $\rho_n \Rightarrow \rho$  and  $\lim_n \alpha_n = \alpha$  where for all  $n$ ,  $\rho_n(\mathbb{R}) = 1$  and  $\rho_n(\alpha_n) < 1$  with  $\alpha_n \in \mathcal{B}(\mathbb{R})$ ,  $\rho$  is a probability measure on  $\mathcal{B}(\overline{\mathbb{R}})$  with  $\rho(\{-\infty\}) = 0$ , and  $\alpha \in \mathcal{B}(\mathbb{R})$ . If  $\lim_n \rho_n(\alpha_n)$  is a probability,  $\lim_n [\rho_n(\alpha_n)]$  may not necessarily be a probability. There is a subtle difference between  $\lim_n \rho_n(\alpha_n)$  and  $\lim_n [\rho_n(\alpha_n)]$ . As shown in Example 4 and Example 11, if  $\rho(\mathbb{R}) = 0$  and if  $\lim_n [\rho_n(\alpha_n)] > 0$  exists, then  $\lim_n [\rho_n(\alpha_n)]$  is not a probability. On the other hand, if  $\rho(\mathbb{R}) = 1$ , then  $\lim_n [\rho_n(\alpha_n)]$  is also a probability and coincides with  $\lim_n \rho_n(\alpha_n)$ . A limiting probability cannot always be expressed as the limit of a real sequence. Taking limit for a real sequence is not necessarily a closed operation with respect to probabilities. Nevertheless, the definition of the limit of a real sequence remains unchanged, even if the limit is not a probability. We shall prove the above claims in this section. The case  $0 < \rho(\mathbb{R}) < 1$  may be reduced to  $\rho'(\mathbb{R}) = 1$  where  $\rho' = \rho/\rho(\mathbb{R})$  such that  $\rho'_n \Rightarrow \rho'$  and  $\rho'_n = \rho_n/\rho_n(\mathbb{R})$  [7].

To clarify the ambiguity about  $\lim_n \rho_n(\alpha_n)$ , we introduce a mapping. Let  $\psi$  be a probability measure on  $\mathcal{B}(\overline{\mathbb{R}})$ , and  $\eta$  a measurable set in  $\mathcal{B}(\mathbb{R})$ . Denote by  $\mathcal{P}$  a mapping, which maps  $(\psi, \eta)$  to  $\psi(\eta)$ , i.e.,  $\mathcal{P}(\psi, \eta) = \psi(\eta)$ . With  $\mathcal{P}$ ,  $\lim_n [\rho_n(\alpha_n)]$  can be expressed by  $\lim_n \mathcal{P}(\rho_n, \alpha_n)$ . If  $\lim_n \rho_n(\alpha_n)$  is a probability, we use  $[\lim_n \rho_n(\alpha_n)]_m$  and  $[\lim_n \rho_n(\alpha_n)]_s$  to represent the respective probability measure and measurable set which produce  $\lim_n \rho_n(\alpha_n)$ .

**Remark 1** In general, the argument of  $[\cdot]_m$  and  $[\cdot]_s$  corresponds to a procedure for calculating a probability. The procedure may simply be a specific probability assignment such as  $\psi(\eta)$ , or is given by an expression for computing a limiting probability, which can eventually be reduced to the form of  $\rho(\alpha)$ , where  $\rho$  is the measure of the limiting probability and  $\alpha$  the associated  $\rho$ -measurable set, determined respectively by weak convergence of the sequence of involved probability measures  $\rho_n$  and the corresponding measurable sets  $\alpha_n$  in the context. Procedures for calculating limiting probabilities may not



necessarily be the same as those for calculating limits of real sequences. A procedure for calculating the limit of a real sequence may not necessarily yield a limiting probability as shown by Example 11.  $\square$

**Lemma 1** Suppose  $\rho_n \Rightarrow \rho$  with  $\rho(\{-\infty\}) = 0$ ,  $\lim_n \alpha_n = \alpha$ , and  $\rho_n(\alpha_n) < 1$  with  $\alpha_n \in \mathcal{B}(\mathbb{R})$  and  $\rho_n(\mathbb{R}) = 1$  for all  $n$ .

(a) If  $\lim_n \rho_n(\alpha_n)$  is a probability,

$$\lim_{n \rightarrow \infty} \rho_n(\alpha_n) = \mathcal{P}(\rho, \alpha).$$

(b) If both  $\lim_n \rho_n(\alpha_n)$  and  $\lim_n \mathcal{P}(\rho_n, \alpha_n)$  are probabilities,

$$\lim_{n \rightarrow \infty} \rho_n(\alpha_n) = \mathcal{P}(\rho, \alpha) = \lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, \alpha_n).$$

**Proof:** (a) If  $\lim_n \rho_n(\alpha_n)$  is a probability, then from  $\rho_n \Rightarrow \rho$  and  $\lim_n \alpha_n = \alpha$  we have

$$\left[ \lim_{n \rightarrow \infty} \rho_n(\alpha_n) \right]_m = \rho \text{ and } \left[ \lim_{n \rightarrow \infty} \rho_n(\alpha_n) \right]_s = \alpha.$$

Immediately from the definitions of  $[\cdot]_m$ ,  $[\cdot]_s$  and  $\mathcal{P}$ ,

$$\lim_{n \rightarrow \infty} \rho_n(\alpha_n) = \left[ \lim_{n \rightarrow \infty} \rho_n(\alpha_n) \right]_m \left( \left[ \lim_{n \rightarrow \infty} \rho_n(\alpha_n) \right]_s \right) = \rho(\alpha) = \mathcal{P}(\rho, \alpha).$$

(b) If  $\lim_n \rho_n(\alpha_n)$  and  $\lim_n \mathcal{P}(\rho_n, \alpha_n)$  are both probabilities, we also have  $[\lim_n \mathcal{P}(\rho_n, \alpha_n)]_m = \rho$ ,  $[\lim_n \mathcal{P}(\rho_n, \alpha_n)]_s = \alpha$ , and

$$\lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, \alpha_n) = \left[ \lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, \alpha_n) \right]_m \left( \left[ \lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, \alpha_n) \right]_s \right) = \rho(\alpha) = \mathcal{P}(\rho, \alpha)$$

which together with (a) proves (b).  $\square$

Thus, if  $\lim_n \rho_n(\alpha_n)$  is a probability, a general expression, or so to speak, a synonym of  $\lim_n \rho_n(\alpha_n)$  is  $\mathcal{P}(\rho, \alpha)$ .

**Remark 2** By Lemma 1, given  $\rho_n \Rightarrow \rho$  and  $\lim_n \alpha_n = \alpha$ , if both  $\lim_n \rho_n(\alpha_n)$  and  $\lim_n [\rho_n(\alpha_n)]$  are probabilities, then they can be expressed respectively as  $\mathcal{P}(\rho, \alpha)$  and  $\lim_n \mathcal{P}(\rho_n, \alpha_n)$ , and the order of applying  $\mathcal{P}$  and taking limit is interchangeable. Labeled by a single index variable  $n$ ,  $\lim_n \rho_n(\alpha_n)$  is not a double limit. A double limit involves a double sequence labeled by two independent index variables (e.g.,  $(a_{ij})_{i \geq 1, j \geq 1}$ ). Issues about a double limit, such as whether different orders of taking limit will yield the same result, are irrelevant to  $\lim_n \rho_n(\alpha_n)$ .  $\square$

The following is a variant of Lemma 1 in some specific settings with  $\alpha_n = I(x_n) = (-\infty, x_n]$ , where  $(x_n)_{n \geq 1}$  is a real sequence, such that  $\rho_n(I(x_n)) < 1$  for all  $n$ , and as  $n \rightarrow \infty$ ,  $x_n \uparrow \infty$  or  $x_n \rightarrow x \in \mathbb{R}$ .

**Lemma 2** Assume that  $\lim_n \rho_n(I(x_n))$  is a probability,  $\rho_n \Rightarrow \rho$  with  $\rho(\{-\infty\}) = 0$ , and  $\rho_n(I(x_n)) < 1$  for all  $n$ .

(a) If  $x_n \rightarrow x \in \mathbb{R}$  as  $n \rightarrow \infty$  where  $\rho(\{x\}) = 0$ , then  $\lim_n \mathcal{P}(\rho_n, I(x_n))$  is a probability, and

$$\lim_{n \rightarrow \infty} \rho_n(I(x_n)) = \mathcal{P}(\rho, I(x)) = \lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, I(x_n))$$

where  $I(x) = \lim_n I(x_n) = (-\infty, x]$ .

(b) Suppose  $x_n \uparrow \infty$  as  $n \rightarrow \infty$ . Then

$$\lim_{n \rightarrow \infty} \rho_n(I(x_n)) = \mathcal{P}(\rho, \mathbb{R}) \leq 1.$$

If  $(\rho_n)_{n \geq 1}$  is tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  and if  $\lim_n \mathcal{P}(\rho_n, I(x_n))$  is a probability,

$$\lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, I(x_n)) = \mathcal{P}(\rho, \mathbb{R}) = 1.$$

If  $(\rho_n)_{n \geq 1}$  is not tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , then  $\lim_n \mathcal{P}(\rho_n, I(x_n))$  (if the limit exists) is not a probability, and

$$\mathcal{P}(\rho, \mathbb{R}) = 1 - \rho(\{\infty\}) < 1.$$

**Proof:** (a) Since  $\rho_n \Rightarrow \rho$  and since  $x_n \rightarrow x \in \mathbb{R}$ , given  $\epsilon > 0$ , there exists an  $N_\epsilon$  such that for all  $n \geq N_\epsilon$ ,

$$|\rho_n(I(x_n)) - \rho(I(x))| \leq |\rho_n(I(x_n)) - \rho_n(I(x))| + |\rho_n(I(x)) - \rho(I(x))| < \epsilon/2 + \epsilon/2 = \epsilon.$$

So

$$\lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, I(x_n)) = \rho(I(x)) = \mathcal{P}(\rho, I(x)).$$

On the other hand,

$$\lim_{n \rightarrow \infty} \rho_n(I(x_n)) = \left[ \lim_{n \rightarrow \infty} \rho_n(I(x_n)) \right]_m \left( \left[ \lim_{n \rightarrow \infty} \rho_n(I(x_n)) \right]_s \right) = \rho(I(x)).$$

This completes the proof of (a).

(b) Since  $\lim_n I(x_n) = \cup_{n \geq 1} I(x_n) = \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} \rho_n(I(x_n)) = \left[ \lim_{n \rightarrow \infty} \rho_n(I(x_n)) \right]_m \left( \left[ \lim_{n \rightarrow \infty} \rho_n(I(x_n)) \right]_s \right) = \rho(\mathbb{R}) = \mathcal{P}(\rho, \mathbb{R}) \leq 1.$$

If  $(\rho_n)_{n \geq 1}$  is tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , no mass “escapes to infinity” in probability assignment given by  $\rho$ , and hence  $\mathcal{P}(\rho, \mathbb{R}) = 1$ . Moreover, for each  $\epsilon > 0$ , there is an interval  $I_{b(\epsilon)} = (-\infty, b(\epsilon)]$ , such that

$$\sup_{n \geq 1} \rho_n(I_{b(\epsilon)}^c) = \sup_{n \geq 1} \mathcal{P}(\rho_n, I_{b(\epsilon)}^c) < \epsilon$$

where  $I_{b(\epsilon)}^c = \mathbb{R} \setminus I_{b(\epsilon)}$ . For all  $n \geq N_\epsilon$  where  $N_\epsilon = \min\{n \in \mathbb{N} : x_n > b(\epsilon)\}$ ,  $I(x_n) \supset I_{b(\epsilon)}$ ,  $I^c(x_n) (= \mathbb{R} \setminus I(x_n)) \subset I_{b(\epsilon)}^c$ , and hence

$$\mathcal{P}(\rho_n, I^c(x_n)) \leq \mathcal{P}(\rho_n, I_{b(\epsilon)}^c) \leq \sup_{n \geq 1} \mathcal{P}(\rho_n, I_{b(\epsilon)}^c) < \epsilon.$$

In other words,  $\mathcal{P}(\rho_n, I^c(x_n)) \rightarrow 0$  as  $n \rightarrow \infty$ . Consequently,  $\mathcal{P}(\rho_n, I(x_n)) \rightarrow 1$ .

Now assume that  $(\rho_n)_{n \geq 1}$  is not tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . In probability assignment given by  $\rho_n$  for each  $n$ , mass must be conserved on the extended real line. Consequently

$$\rho_n(I(x_n)) = \rho_n(\overline{\mathbb{R}}) - \rho_n(I^c(x_n)) - \rho_n(\{\infty\}). \quad (2)$$

Note  $\rho_n(I^c(x_n)) + \rho_n(\{\infty\}) = \rho_n(\overline{\mathbb{R}} \setminus I(x_n))$ . In (2),  $\rho_n(\{\infty\}) = 0$  and  $\rho_n(\overline{\mathbb{R}}) = 1$ , since by assumption, for each  $n$ ,  $\rho_n(\mathbb{R}) = 1$ . Thus, with  $\mathcal{P}$ , we can write (2) as

$$\mathcal{P}(\rho_n, I(x_n)) = 1 - \mathcal{P}(\rho_n, I^c(x_n)).$$

Taking limit on both sides of this equation yields (if the limits exist)

$$\lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, I(x_n)) = 1 - \lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, I^c(x_n)). \quad (3)$$

If  $\lim_n \mathcal{P}(\rho_n, I(x_n))$  is a probability, so is  $\lim_n \mathcal{P}(\rho_n, I^c(x_n))$ , such that

$$\begin{aligned} \left[ \lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, I(x_n)) \right]_s &= \lim_{n \rightarrow \infty} I(x_n) = \mathbb{R} \\ \left[ \lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, I^c(x_n)) \right]_s &= \lim_{n \rightarrow \infty} I^c(x_n) = \bigcap_{n \geq 1} I^c(x_n) = \emptyset \end{aligned}$$

and (3) implies

$$\left[ \lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, I(x_n)) \right]_m = \left[ \lim_{n \rightarrow \infty} \mathcal{P}(\rho_n, I^c(x_n)) \right]_m = \rho$$

with  $\rho(\mathbb{R} \cup \emptyset) = \rho(\mathbb{R}) = 1$ . However, given that  $(\rho_n)_{n \geq 1}$  is not tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , we have  $\rho(\{\infty\}) > 0$ . Consequently,  $\rho(\mathbb{R}) = 1$  implies

$$\rho(\mathbb{R}) + \rho(\{\infty\}) = \rho(\overline{\mathbb{R}}) > 1$$

and hence violates the constraint imposed by the axioms of probability. As a probability measure on  $\mathcal{B}(\overline{\mathbb{R}})$ ,  $\rho$  assigns unit mass to  $\overline{\mathbb{R}}$ . So  $\lim_n \mathcal{P}(\rho_n, I(x_n))$  is not a probability. Similarly, from mass conservation on  $\overline{\mathbb{R}}$  in probability assignment given by  $\rho$ ,

$$\rho(\mathbb{R}) = \rho(\overline{\mathbb{R}}) - \rho(\{\infty\}) = \mathcal{P}(\rho, \mathbb{R}) = 1 - \rho(\{\infty\}) < 1$$

follows immediately.  $\square$

## 8 The Inconsistency

Now we demonstrate the inconsistency of measure-theoretic probability. Let  $\overline{\mathbb{R}}$  be equipped with a topology induced by a fixed metric (e.g., see Example 6). Let  $\lambda, \mu, \lambda_n, \mu_n, n = 1, 2, \dots$  be probability measures on  $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ , such that  $\lambda, \lambda_1, \lambda_2, \dots$  are distributions of i.i.d. random variables  $X, X_1, X_2, \dots$ , respectively, as given in Example 3, and  $\mu_n$  are distributions of

$$Z_n = \max\{i : X_i = Y_n, i \in \{1, 2, \dots, n\}\} \quad (4)$$

where

$$Y_n = \max\{X_i : i \in \{1, 2, \dots, n\}\}. \quad (5)$$

By (4),  $Z_n, n = 1, 2, \dots$  form a non-decreasing random sequence  $(Z_n)_{n \geq 1}$ . As we can readily see, for each  $n$ , the probability mass function of  $Z_n$  is given by (1) in Example 4. As  $n \rightarrow \infty$ ,  $Z_n$  converges in distribution to an extended random variable  $Z$ . Equivalently,  $\mu_n \Rightarrow \mu$ , i.e.,  $\mu_n$  converges weakly to  $\mu$  induced by  $Z$ . As a probability measure on  $\mathcal{B}(\overline{\mathbb{R}})$ ,  $\mu$  is the Dirac measure concentrated at  $\{\infty\}$ , i.e.,  $\mu(\{\infty\}) = 1$ . However, when restricted to  $\mathcal{B}(\mathbb{R})$ ,  $\mu$  is a subprobability measure with  $\mu(\mathbb{R}) = 0$ . As shown in Example 3 and Example 4,  $(\lambda_n)_{n \geq 1}$  is tight on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  but  $(\mu_n)_{n \geq 1}$  is not (see Sections 4 and 5). By (4) and (5),

$$Z_n \leq n \quad (6)$$

and

$$X_{Z_n} = Y_n. \quad (7)$$

For the random sequences  $(X_n)_{n \geq 1}$  and  $(Y_n)_{n \geq 1}$ , a probability measure  $\mathbb{P}$  can be defined on the measurable space  $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$  where  $T = \{1, 2, \dots\}$ . This probability measure  $\mathbb{P}$  can also express probabilities given by the distribution  $\mu_n$  induced by  $Z_n$  for each  $n$ . However, as shown in Section 3, no probability measure can be defined on  $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$  to capture the impact of the infinite limit of  $(Z_n)_{n \geq 1}$  on limiting probabilities given by  $\mu$ .

In contrast, for the notion of weak convergence, probability spaces which are the domains of the involved random variables can all be distinct. The domain spaces are not essential, and actually remain offstage [4, 7]. So we can focus on the probability measures and their weak limits, as long as the range space and the topology on it are the same for all the random variables involved [4]. The weak limit of  $\lambda_n$  is trivially  $\lambda$ , i.e.,  $\lambda_n \Rightarrow \lambda$  with  $\lambda(\{0\}) = 1/2$ . However, the following theorem is also provable.

**Theorem 1** *For  $\lambda_n$  given in Example 3,  $\lambda_n \Rightarrow \lambda$  with  $\lambda(\{0\}) = \mu(\mathbb{R}) = 0$ .*

**Proof:** By (6),

$$\lambda_n(\{0\}) = \mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 0, Z_n < n) + \mathbb{P}(X_n = 0, Z_n = n). \quad (8)$$

Consider the limiting probabilities on both sides of (8). We first show  $\lim_n \mathbb{P}(X_n = 0, Z_n = n) = 0$ .

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}(X_n = 0, Z_n = n) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(X_{Z_n} = 0, Z_n = n) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(Y_n = 0, Z_n = n) \text{ (by (7))} \\ &= 0. \end{aligned} \quad (9)$$

The last equality follows from  $\mathbb{P}(\lim_{n \rightarrow \infty} Y_n = 1) = 1$ ,

$$\{Y_n = 0, Z_n = n\} \subset \{Y_n = 0\}, \quad \mathbb{P}(Y_n = 0, Z_n = n) \leq \mathbb{P}(Y_n = 0)$$

and

$$\lim_{n \rightarrow \infty} \mathbb{P}(Y_n = 0) = \mathbb{P}\left(\bigcap_{n \geq 1} \{Y_n = 0\}\right) \text{ (since } \{Y_n = 0\} \supset \{Y_{n+1} = 0\} \text{ as implied by (5))}.$$

Similarly, from (8), (9), and

$$\{X_n = 0, Z_n < n\} = \{Z_n < n\}$$

we have

$$\mathbb{P}\{X_n = 0, Z_n < n\} = \mathbb{P}\{Z_n < n\} = \mu_n(I(n-1))$$

and

$$\lambda(\{0\}) = \lim_{n \rightarrow \infty} \lambda_n(\{0\}) = \lim_{n \rightarrow \infty} \mathbb{P}(X_n = 0, Z_n < n) = \lim_{n \rightarrow \infty} \mu_n(I(n-1)).$$

Since  $\lambda(\{0\})$  is a probability, so is  $\lim_n \mu_n(I(n-1))$ , and

$$\lambda(\{0\}) = \lim_{n \rightarrow \infty} \mu_n(I(n-1)) = \mathcal{P}(\mu, \mathbb{R}) = \mu(\mathbb{R}) = 0$$

follows from Lemma 2(b) with  $x_n = n-1$ ,  $\rho_n = \mu_n$  and  $\rho(\{\infty\}) = \mu(\{\infty\}) = 1$ .  $\square$

**Remark 3** By definition, “ $1/2$ ” in the probability mass function “ $\lambda(\{0\}) = 1/2$ ” (see Example 3) represents a (numerical) probability. In contrast, as explained in Example 4 and Example 11, “ $1/2$ ” in “ $\lim_n [\mu(I(n-1))] = \lim_n \mathcal{P}(\mu_n, I(n-1)) = 1/2$ ” is not a probability. As we have discussed in Sections 6 and 7, without representing the value of a quantity in the physical world, a real number is only a point in a set formulated by a system of axioms. It makes no sense at all to treat a probability, which is a quantified attribute of an event in the physical world, as a point of a set, which is merely a mathematical (or metaphysical) entity still considered questionable by some mathematicians even nowadays [5, 9]. Since  $\lambda(\{0\}) = \lim_n \mu(I(n-1))$  as shown in the proof of Theorem 1, and since  $\lambda(\{0\})$  is a probability,  $\lim_n \mu(I(n-1))$  must be a probability, and this probability is  $\mathcal{P}(\mu, \mathbb{R}) = \mu(\mathbb{R}) = 0$  by Lemma 2. Since “ $1/2$ ” in “ $\lim_n [\mu(I(n-1))] = \lim_n \mathcal{P}(\mu_n, I(n-1)) = 1/2$ ” is a real number, i.e., a point on the abstract real line  $\mathbb{R}$  but not a probability, it is meaningless to equate  $\lambda(\{0\})$  with  $\lim_n [\mu(I(n-1))] = \lim_n \mathcal{P}(\mu_n, I(n-1))$  (see also Example 8 and Example 9).  $\square$

## 9 Concluding Remarks

In general, the presupposition for questions of identity (expressed by “ $=$ ”) in a theory is that the two sides inquired about both belong to the same category [2]. An expression “ $x = y$ ” where  $x$  and  $y$  are in the same category but numerically different is erroneous only when the theory is proved to be contradiction-free as requested by Hilbert. Unfortunately, according to Gödel’s theorems, measure-theoretic probability is one of the theories for which such a proof does not exist. Nevertheless, if one side of “ $=$ ” is a probability, then the other side must also be a probability. This requirement is necessary to prevent confusion of probabilities and arbitrarily calculated numbers which are not eligible to serve as probabilities. Whether a number is eligible to serve as a probability is determined by whether the number is the result of applying a probability measure to a measurable set in the context. In particular, when we calculate limiting probabilities, the two sides of “ $=$ ” must both be limiting probabilities. Since procedures for calculating limits of real sequences may not necessarily yield limiting probabilities, the limit of a real sequence which is not a probability is irrelevant to whatever implied by calculations of limiting probabilities.

Consider  $\lim_n \mu_n(I(n-1))$  (see the proof of Theorem 1). The probability measures  $\mu_n$  and the associated measurable sets  $I(n-1)$  approach their respective limits simultaneously. This implies two different orders of operations for calculating  $\lim_n \mu_n(I(n-1))$ :

- (a) taking the limits of the probability measures  $\mu_n$  and the measurable sets  $I(n-1)$  simultaneously first, then applying the limiting measure  $\mu$  to the limiting measurable set  $\mathbb{R}$ , i.e.,

$$\lim_{n \rightarrow \infty} \mu_n(I(n-1)) = \mathcal{P}(\mu, \mathbb{R}) = \mu(\mathbb{R})$$

- (b) applying the measures  $\mu_n$  to the corresponding measurable sets  $I(n-1)$  to obtain a sequence  $(\mu_n(I(n-1)))_{n \geq 1}$  of numbers first, then taking the limit of the real sequence, i.e.,

$$\lim_{n \rightarrow \infty} \mu_n(I(n-1)) = \lim_{n \rightarrow \infty} \mathcal{P}(\mu_n, I(n-1)).$$

Order (a) produces a probability, but order (b) yields a number which is not a probability (see Example 11). So the meaning of  $\lim_n \mu_n(I(n-1))$  is ambiguous.

Since  $\lim_n \mu_n(I(n-1))$  can only be written as a “linear” string of symbols, such a “linear” way of writing might result in an impression as if (b) were the only option for calculating  $\lim_n \mu_n(I(n-1))$ . However, when standing in the relation “ $=$ ” to a probability (i.e.,  $\lambda(\{0\})$ ),  $\lim_n \mu_n(I(n-1))$  has to be

calculated according to (a) rather than (b) to yield a probability as well, for the two sides of “=” must both be probabilities, even if they are numerically different. As we have demonstrated, “two numerically different probabilities equal each other” is a provable statement, which shows that measure-theoretic probability is inconsistent.

The revealed contradiction of measure-theoretic probability does not exist in the physical world. It appears only in the theory. Practical applications of measure-theoretic probability will not be affected by the inconsistency as long as ideal events in the theory are not mistaken for real events in the physical world. On the other hand, the inconsistency of measure-theoretic probability cannot be explained away, and must be resolved. Constructive mathematics (e.g., see [5, 9]) can avoid such inconsistency. There is no contradiction reported in constructive mathematics.

## References

- [1] P. Benacerraf and H. Putnam eds. **Philosophy of Mathematics: Selected Readings**, Cambridge University Press, 2nd ed., New York, 1983.
- [2] P. Benacerraf, What numbers could not be, *Philosophical Review*, 74(1965): 47-73.
- [3] P. Billingsley, **Probability and Measure**, John Wiley & Sons, New York, 3ed., 1995.
- [4] P. Billingsley, **Convergence of Probability Measures**, John Wiley & Sons, New York, 2nd ed., 1999.
- [5] D. Bridges and F. Richman, **Varieties of Constructive Mathematics**, London Mathematical Society Lecture Notes Series, vol. 97, Cambridge University Press, Cambridge, 1987.
- [6] J. Czyż, **Paradoxes of Measures and Dimensions Originating in Felix Hausdorff’s Ideas**, World Scientific, Singapore, 1994.
- [7] M. Loève, **Probability Theory**, vol. 1, Springer-Verlag, New York, 4th ed., 1977.
- [8] D. Pollard, **Convergence of Stochastic Processes**, Springer-Verlag, New York, 1984.
- [9] M. Rosenblatt, Ed., **Errett Bishop: Reflections on Him and His Research**, Contemporary Mathematics, vol. 39, AMS, Providence, Rhode Island, 1985.
- [10] G. J. Székely, **Paradoxes in Probability Theory and Mathematical Statistics**, D. Reidel, Dordrecht, 2001.